

# Spot cycle reconstruction: an empirical tool

## Application to the sunspot cycle

A. R. G. Santos<sup>1,2,3,4</sup>, M. S. Cunha<sup>1,2,3</sup>, P. P. Avelino<sup>1,2,3</sup>, and T. L. Campante<sup>4</sup>

<sup>1</sup> Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, 4150-762 Porto, Portugal  
e-mail: asantos@astro.up.pt

<sup>2</sup> Centro de Astrofísica da Universidade do Porto, Rua das Estrelas, 4150-762 Porto, Portugal

<sup>3</sup> Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

<sup>4</sup> School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Received November 7, 2014 / Accepted June 4, 2015

### ABSTRACT

**Context.** The increasing interest in understanding stellar magnetic activity cycles is a strong motivation for the development of parameterized starspot models which can be constrained observationally.

**Aims.** In this work we develop an empirical tool for the stochastic reconstruction of sunspot cycles, using the average solar properties as a reference.

**Methods.** The synthetic sunspot cycle is compared with the sunspot data extracted from the National Geophysical Data Center, in particular using the Kolmogorov-Smirnov test. This tool yields synthetic spot group records, including date, area, latitude, longitude, rotation rate of the solar surface at the group's latitude, and an identification number.

**Results.** Comparison of the stochastic reconstructions with the daily sunspot records confirms that our empirical model is able to successfully reproduce the main properties of the solar sunspot cycle. As a by-product of this work, we show that the Gnevyshev-Waldmeier rule, which describes the spots' area-lifetime relation, is not adequate for small groups and we propose an effective correction to that relation which leads to a closer agreement between the synthetic sunspot cycle and the observations.

**Key words.** Sun: activity – Sun: magnetic fields – Stars: activity – Stars: oscillations – Sunspots – Starspots

### 1. Introduction

In the Sun, the level of magnetic activity varies over time showing a periodic behaviour known as the solar cycle. The most direct evidence for this is the 11-year variation in the number of sunspots observed at the solar photosphere. As the cycle proceeds, the sunspot formation zone gradually migrates towards the equator until the next minimum is reached and a new cycle begins. At that point, the polarity of the magnetic field reverses and returns to the original state at the end of the second 11-year cycle, thus completing a 22-year Hale cycle (the sunspot polarity law - Hale et al. 1919; Hale & Nicholson 1925).

Evidence for the presence of activity cycles in other solar-like stars, including exoplanet hosts, has also been accumulating over the past years. Starspots cannot be observed directly at the surface of distant stars, but can be detected indirectly through the effects they induce, such as the strong emission at the centre of the Ca II H and K lines. By monitoring this emission, it is possible to infer the rotation period of the star and the period of its magnetic cycle (e.g. Wilson 1978; Duncan et al. 1991; Gray & Baliunas 1995; Saar & Brandenburg 1999; Cincunegui et al. 2007; Hall et al. 2007; Metcalfe et al. 2010, 2013). Moreover, since spots are darker than the average stellar surface they can also be detected through the inspection of the photometric light curves of stars observed by CoRoT (Baglin et al. 2006) and Kepler (Borucki et al. 2010; Koch et al. 2010) space missions. In an active star with starspots crossing the visible disk, the light curve shows a quasi-periodic modulation which results from the

combination of the effects of stellar rotation and magnetic activity (e.g. Mosser et al. 2009; Mathur et al. 2010; García et al. 2010b; Ballot et al. 2011; Campante et al. 2011). That modulation is more significant in periods of maximum activity, making the light curves a possible starspot proxy. In addition, through the wavelet analysis of the light curves it might be possible to obtain information about the time evolution of the starspots at the stellar surface and, in turn, derive the period of the activity cycle and the rotation period of the star (e.g. Campante 2012; Mathur et al. 2014; García et al. 2014; Bravo et al. 2014). Finally, stellar activity cycles may also be detected through their impact on stellar oscillations because the magnetic activity affects the wave propagation, inducing changes to the oscillation frequencies, amplitudes, and line widths, which are thus found to vary in phase with other sun-/starspot proxies (e.g. Woodard & Noyes 1985; Libbrecht & Woodard 1990; Chaplin et al. 2004; Metcalfe et al. 2007; García et al. 2010a; Tripathy et al. 2011).

Starspots simulations can be used to study how the activity-induced variability in the light curve and in the oscillation properties depends on the characteristics of the activity cycle. These simulations are important in other astrophysical contexts, such as in the quest for exoplanets, where they can be used to design new strategies to reduce the signatures induced by stellar activity on the radial velocity and transit observations (e.g. Pont et al. 2008; Czesla et al. 2009; Figueira et al. 2010; Dumusque et al. 2011; Oshagh et al. 2013, 2014). With the above in mind, in this work we develop a parameterized model aimed at reproducing

the main properties of the sunspot cycle that can also be applied to reproducing the activity cycles of other stars. In order to build our empirical model, we use as inputs a number of properties of the observed sunspot cycle. These properties are summarized in section 2 and their implementation in our model is discussed in section 3. The results obtained with this tool and their comparison with the solar data extracted from the National Geophysical Data Center (NOAA/NGDC) are presented in section 4. Finally, in section 5 we draw our main conclusions.

## 2. Properties of the solar cycle

In order to reproduce the sunspot cycle, a number of important observational constraints must be considered. In what follows we review key observational properties of the sunspot cycle that will be used in our model, namely: the number of sunspot groups and its dependence on the phase of the cycle; the sunspots' areas and their relation with the sunspots' lifetimes; the formation latitude of the sunspots and the width of the formation region; the differential rotation of the solar surface.

The number of sunspots and sunspot groups varies over the solar cycle. Its evolution is asymmetric: the rising phase of the cycle is faster than the declining towards the next minimum. Different authors have used different functions to describe the asymmetric shape of the sunspot cycle (e.g. Setiawati & Panofsky 1938; Elling & Schwentek 1992; Sabarinath & Anilkumar 2008; Volobuev 2009; Du 2011). In particular, Hathaway, Wilson, & Reichmann (1994) found that the observed number of sunspots is nicely fitted by the following function of time,

$$N_S(t) = \frac{a_1(t - t_0)^3}{\exp((t - t_0)^2/b_1^2) - c_1}, \quad (1)$$

where  $t_0$  is the starting time (about four months prior to the minimum for an average cycle (Hathaway 2010)),  $a_1$  is the amplitude,  $b_1$  is related to the size of the cycle, and  $c_1$  is related to the asymmetry of the cycle.

The asymmetric shape of the solar cycle is also evident in the temporal variation of the sunspot areas. Bogdan et al. (1988) were the first to notice that the accumulated umbral areas distribution (derived from daily records by counting each spot as many times as the number of days it remains visible) can be described by a log-normal distribution. Since the ratio between the umbral area and the total area of the spot does not depend significantly on the spot size (e.g. Brandt et al. 1990; Solanki 2003; Vaquero et al. 2005; Kiess et al. 2014), the accumulated spot areas also follow a log-normal distribution. This finding was confirmed by later studies (e.g. Baumann & Solanki 2005; Hathaway & Choudhary 2008; Kiess et al. 2014). Moreover, Baumann & Solanki (2005) have shown that the log-normal distribution also nicely fits the observed distributions for the instantaneous area and for the maximum area of sunspots. However, according to Jiang et al. (2011), the area distribution for groups smaller than 60 MSH (millionth of the solar hemisphere) is better described by a power law.

During its life, a given sunspot group grows until it reaches a maximum area and then decays. The growth ( $\Psi$ ) and decay ( $\Gamma$ ) rates, i.e. the time derivative of the group's area during each of these phases, are found to be dependent on the group areas, the activity cycle, the phase of the cycle, and the latitude (e.g. Moreno-Insertis & Vázquez 1988; Howard 1992; Petrovay & van Driel-Gesztelyi 1997; Hathaway & Choudhary 2008; Javaraiah 2012). Small groups grow faster than they decay, while the growth rates of large groups are smaller than their decay rates

(Howard 1992). Hathaway & Choudhary (2008) found a linear relation between the decay rates and the group area, but the erosion model proposed earlier by Petrovay & van Driel-Gesztelyi (1997) indicates a non-linear relation of the type  $\Gamma \propto A^{0.5}$ , where  $A$  is the group area at a given time. More recently, Javaraiah (2012) suggested that the relation between the decay rates and areas could be better described by a power law of the form  $\Gamma = \exp(\gamma_1)A^{\gamma_2}$ . The constant  $\gamma_2$  was found to vary from  $\sim 0.45$  to  $\sim 0.70$ , when considering individual cycles and different phases of each cycle (with  $A$  expressed in MSH). When assuming the sunspot data from 1874–2011, Javaraiah (2012) found that  $\gamma_1 \sim 0.26$  and  $\gamma_2 \sim 0.613$ . On the other hand, the same study did not produce conclusive results regarding the relation between the group's growth rate and its area.

The areas of sunspots and sunspot groups are also related to their lifetimes, which can range from hours to months, depending on their size. This dependency is described by the Gnevyshev-Waldmeier (GW) rule (Gnevyshev 1938; Waldmeier 1955), according to which,

$$A_m = D_{GW}T. \quad (2)$$

Here,  $A_m$  is the sunspot or sunspot group maximum area (in MSH),  $T$  is the corresponding lifetime, and  $D_{GW}$  is a constant of proportionality (around  $10 \text{ MSH day}^{-1}$ ). The determination of a precise value for  $D_{GW}$  is hampered by the difficulty in measuring the spots' lifetimes due to the nightfall, the solar rotation (lack of observations of the invisible side of the Sun), and limb darkening (e.g. Henwood et al. 2009; Blanter et al. 2006; Solanki 2003). In spite of these difficulties, some studies have been carried out, indicating that  $D_{GW}$  might be larger than first estimated. Petrovay & van Driel-Gesztelyi (1997) found that  $D_{GW}$  is  $10.89 \pm 0.18 \text{ MSH day}^{-1}$  for individual sunspots. More recently, Henwood et al. (2009) studied long-lived sunspot groups and estimated that  $D_{GW} = 11.73 \pm 0.26 \text{ MSH day}^{-1}$ .

As was first reported by Carrington (1863), the sunspot formation latitude also varies periodically with time. At the beginning of a new cycle the first spots appear at latitudes of about  $\pm 40^\circ$ . The succeeding spots form at progressively lower latitudes, being rarely observed within  $\pm 5^\circ$ . At the solar minimum the last spots of the cycle emerge at low latitudes, while spots of the new cycle start to form at high latitudes. This behaviour is known as the Spörer law and it may be seen in the butterfly diagram (or Maunder diagram; Maunder 1904). Despite the presence of short plateaus at intermediate latitudes ( $\sim 10^\circ$  – around maximum; Chang 2012), Hathaway (2011) found that the drift of the sunspot zones follows an exponential function, where the average latitude,  $L_S$ , is given by

$$L_S(t) = L_0 \exp\left(-\frac{t - t_0}{7.5}\right), \quad (3)$$

where  $L_0$  is the mean latitude at the time  $t_0$ , and  $t$  is expressed in years. By considering the intermediate phases of the cycle where there is no overlap between consecutive cycles, Jiang et al. (2011) verified that the evolution of the average latitude can also be described by a second-order polynomial.

Chang (2012) found that the spatial distribution of the sunspot groups at each time  $t$  is bimodal and that it can be described by a double Gaussian; instead, Ivanov et al. (2011) showed that one single Gaussian describes the data reasonably well. Moreover, the width of the sunspot formation zone,  $\sigma_L$ , also varies over the solar cycle (Gleissberg 1958). According to Miletskii & Ivanov (2009), Ivanov et al. (2011), and Ivanov & Miletskii (2011), this width is a linear function of the activity

level. However, according to Jiang et al. (2011), it can be described in relation to the average sunspot group latitude by a second-order polynomial of the form

$$\frac{\sigma_L}{L_S} = a_\sigma + b_\sigma \frac{t - t_{min}}{P_c} + c_\sigma \left( \frac{t - t_{min}}{P_c} \right)^2, \quad (4)$$

where  $a_\sigma$ ,  $b_\sigma$ , and  $c_\sigma$  are the coefficients of the polynomial,  $t_{min}$  corresponds to the minimum, and  $P_c$  is the period of the cycle.

We note that sunspots are depressed (Wilson depression), thus they move according to the subphotospheric layers, i.e. slightly faster than the solar surface (e.g. Zappalà & Zuccarello 1991; Zuccarello 1993; Abuzeid & Marik 1997; Schou et al. 1998; Kitchatinov 2011). Since the spots' depths decrease as they evolve, younger spots also move faster than the older ones.

Finally, we note that the properties described above vary from cycle to cycle and some are found to be correlated (e.g. Solanki et al. 2002, 2008; Hathaway 2010; Jiang et al. 2011).

### 3. Empirical solar cycle model

The primary goal of this work is to produce a tool capable of reproducing an activity cycle that retains the main observational properties of the solar cycle. To that end, we develop an empirical model to generate sunspot groups as a function of time and gradually adapt the model assumptions and inputs until our goal is reached. To decide whether or not a given assumption/input is a better representation of the observational data than the previous ones we use the Kolmogorov-Smirnov test, which is described in section 4.2.

#### 3.1. Number of sunspot groups

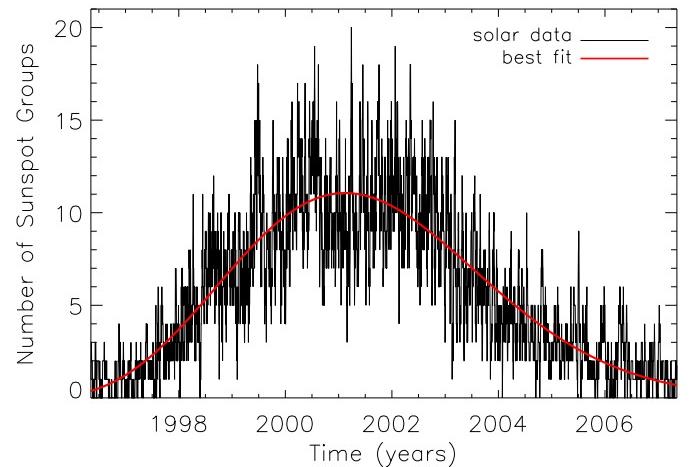
In our model, each sunspot group is generated independently from the others. This could not be assumed if we were considering individual spots, since the formation of spots within the same group is not independent.

At each time step (fixed on one day to be comparable to the daily records of the sunspot data),  $N$  groups are formed. The number of generated groups is randomly determined using a Poisson distribution with a mean value  $N_m$  that depends on time. In the current version of the model,  $N_m$  is taken to be one sixth of  $N_s$ , where the function of time  $N_s$  is derived from a fit of equation (1) to the number of observed sunspot groups for solar cycle 23 (Fig. 1). With this choice for  $N_m$ , we find that the function  $N_s$  derived from the fit to the synthetic data is always in reasonable agreement with that derived from the solar data.

Other ways to determine  $N_s$  from the solar data were explored. The functions used by Du (2011) and Sabarinath & Anilkumar (2008) (the latter takes into account the double peak feature of the solar cycle) led to results that are comparable to those obtained with equation (1).

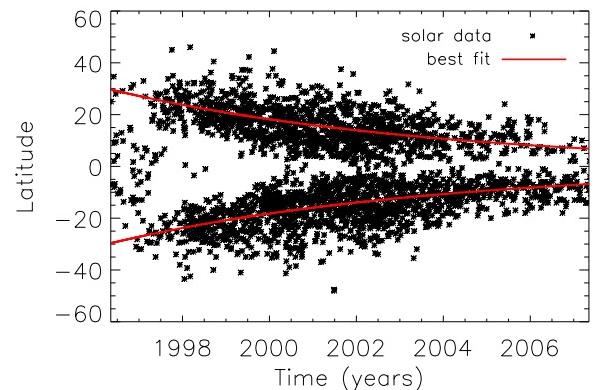
#### 3.2. Sunspot formation zone

In our model, the latitude of each spot,  $L$ , is determined randomly through a single-Gaussian distribution, with a mean latitude  $L_S$  and dispersion  $\sigma_L$ , both dependent on time. For  $L_S$ , we use Hathaway's exponential function given by equation (3) fitted to the northern hemisphere solar data for cycle 23 (Fig. 2), with  $t_0$  fixed at the value found in section 3.1. We opted to use solar data from a single hemisphere because in our model we do not account for the north-south asymmetry. For  $\sigma_L$ , we assume a second-order polynomial (Fig. 3), as suggested by Jiang et al.

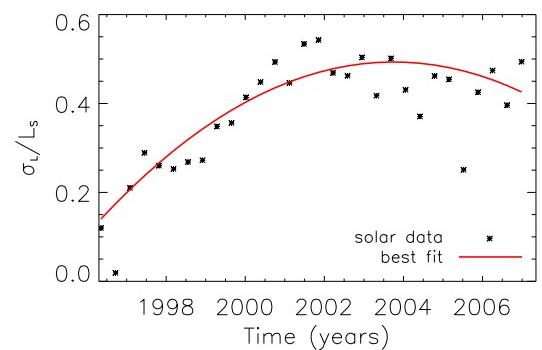


**Fig. 1.** Variation of the number of sunspot groups over the solar cycle 23 (black). The red smooth line corresponds to the fit obtained using the function defined by Hathaway et al. (1994). The observational data were extracted from NOAA.

(2011). Other functions were considered to describe the latitudinal distribution of the sunspot groups (both for  $L_S$  and  $\sigma_L$ ). However, those led to a lower level of agreement between the synthetic and the observational data.



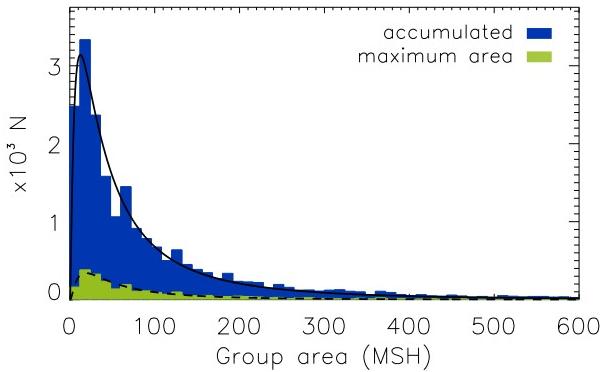
**Fig. 2.** Comparison between the latitudinal distribution of sunspots and the exponential fit of Hathaway (2011). The observational data was extracted from NOAA.



**Fig. 3.** Ratio between the standard deviation of the latitudinal distribution and the corresponding mean latitude (observational data from NOAA). The red solid line corresponds to the best fit.

### 3.3. Sunspot group areas and lifetimes

Having defined the procedure to generate the number of sunspot groups and their position in latitude at each time step, we need to associate a maximum area ( $A_m$ ) to each of them. Although the mean group area varies over the solar cycle (e.g. Jiang et al. 2011), we fix the area distribution. In accordance with the discussion in section 2, in our model the sunspot group maximum areas are drawn from a log-normal distribution whose parameters are obtained from a log-normal fit to the sunspot groups observations for cycle 23, considering each group only once and its maximum area (Fig. 4).



**Fig. 4.** Accumulated (blue) and maximum (green) area distributions for the solar data from NOAA. The black lines are the respective log-normal fits.

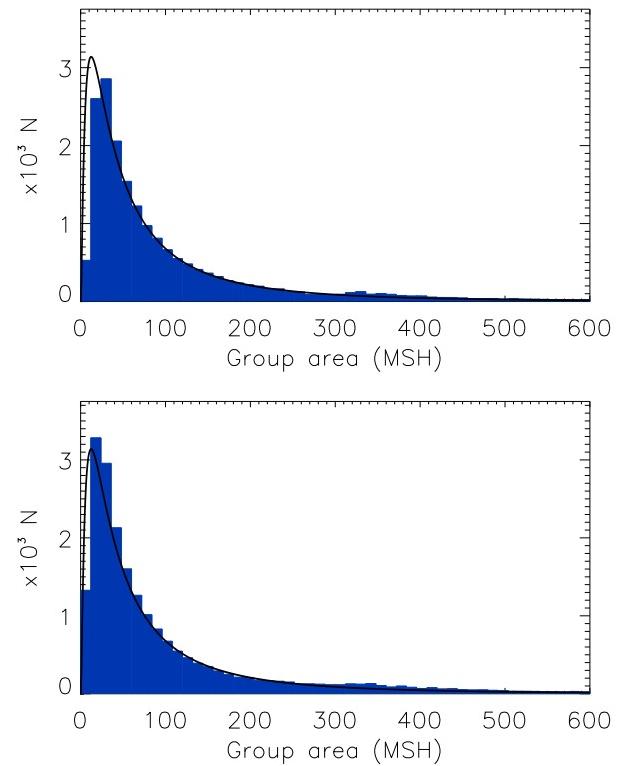
With the sunspot maximum area in hand, we can in principle determine the group's lifetime through the GW rule. However, we have found that when the GW rule is taken for all ranges of areas, the accumulated area distribution obtained with the empirical tool is not in agreement with the observed distribution: the number of small groups in the synthetic distribution is lower than that found in the Sun and the peaks of the two distributions do not coincide. The upper panel of Fig. 5 illustrates this disagreement.

The GW rule is based on the observation of long-lived spot groups (e.g. groups that live longer than  $\sim 20$  days; Henwood et al. 2009). Hence, it is possible that this rule may not be adequate for the smallest groups. In fact, by tracking the small groups in the solar data for cycle 23 and comparing the time they remain visible with the lifetime predicted by the GW rule, it is possible to verify that the GW rule generally underestimates the lifetimes of the small groups. Moreover, there is a significant dispersion around the area-lifetime relation, which for the smallest groups is strongly asymmetric (since lifetimes cannot be negative). With this in mind, we have checked whether increasing the lifetimes for the small groups would improve the agreement between the observed and synthetic accumulated area distributions and found that substituting the GW linear relation by an exponential relation at the lower areas end, our model produces an accumulated area distribution that is in better agreement with the observed one (lower panel of Fig. 5). The modification of the GW rule for groups with areas smaller than 85 MSH used in our model is

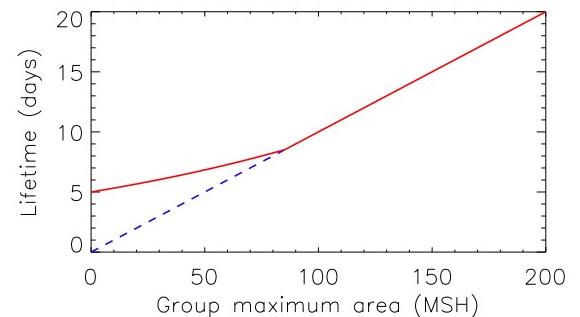
$$T = 5 \exp(6.2591 \times 10^{-3} A_m). \quad (5)$$

Figure 6 shows the current area-lifetime relation used in our empirical model (red line).

<sup>1</sup> The bin for the smallest groups is incomplete, as the sunspot records are limited to areas larger than 10 MSH.



**Fig. 5.** Accumulated area distribution for synthetic sunspot groups when assuming the GW rule for all areas (upper panel) and when using the modified area-lifetime relation (lower panel). The black, solid line is the log-normal fit to the observational data.<sup>1</sup>



**Fig. 6.** The red line shows the area-lifetime relation assumed in the current model while the blue dashed line shows the lifetime predicted for small groups by the GW rule.

Taking the group's maximum area and lifetime, we determine the group's area at a posterior time (while  $A > 0$ ) by applying a decay rate of the form  $\Gamma = \exp(\gamma_1)A^{\gamma_2}$ . We then assume that the period during which the group grows corresponds to the difference between its lifetime and the decay time. In the absence of a well-established relation between the group's growth rate and its area, we opted for a power law to describe that relation, i.e.  $\Psi = \exp(\psi_1)A^{\psi_2}$ . We started by considering the values found by Javaraiah (2012) for  $\gamma_1$  and  $\gamma_2$ , making  $\gamma_1 = \psi_1 = 0.26$  and  $\gamma_2 = \psi_2 = 0.613$ . These parameters were progressively changed until a reasonable agreement between the observations and the synthetic data was reached. The current version of the model considers  $\Psi = \exp(0.17)A^{0.46}$  and  $\Gamma = \exp(0.17)A^{0.47}$ , which is consistent with the fact that large groups have higher decay rates than growth rates. For small groups the growth and decay rates given by the expressions above are essentially the same. How-

ever, since the growth time is taken as the difference between the lifetime and the decay time, in practice the smallest groups do not show a growth phase in the daily records, which might be interpreted as a fast growth, where they reach the maximum area in a time shorter than the interval between consecutive records (one day). We also tested the linear relation found by Hathaway & Choudhary (2008), but we found that decay rates are too high when compared with the lifetime of the group. Assuming the linear relation for the growth and decay rates, the time interval from the first appearance (with  $A \sim 0$ ) to the last appearance (with  $A \sim 0$ ) is much shorter than the lifetime from the modified GW rule.

### 3.4. Sunspot visibility

To reproduce the daily sunspot records of the Sun, we take into account the solar rotation and the fact that spots are observed only when they are on the Sun's visible side. The group's longitude is determined randomly from a uniform distribution between 0 and  $2\pi$ . If this quantity is smaller than  $\pi$ , we consider that the sunspot group is on the visible side, otherwise we consider that it cannot be observed. Taking into account that the rotation velocity of the subphotospheric layers is not very different from that of the solar surface, we assume the following parameterization of the groups' rotation velocity,  $\omega$ , as a function of their latitude,  $L$  (Snodgrass 1983; Snodgrass & Ulrich 1990)

$$\omega(L) = 14.71 - 2.33 \sin^2 L - 1.78 \sin^4 L. \quad (6)$$

Spot groups that emerge on the visible side of the Sun can move towards the invisible side and then eventually become visible again depending on their lifetimes and on the solar angular velocity at the latitude they emerge. Moreover, groups that emerge on the invisible side of the Sun can become visible. Both these facts are taken into account in our model and only groups with an area greater than  $\sim 10$  MSH are considered visible (in analogy to the sunspot data).

A schematic summary of our empirical model is shown in Appendix A.

## 4. Results

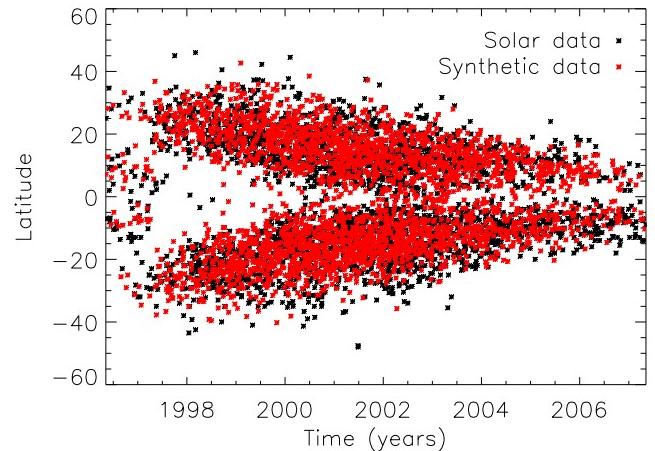
### 4.1. Synthetic data

The synthetic data produced with our empirical model provide information about each generated group. In analogy to what is done in the NOAA databases, the code yields the sunspot group records, which include date, latitude, group area, lifetime, position in longitude, rotation rate of the solar surface at the group's latitude, and an identification number. These records are then used to compare our results with observed data.

A comparison between the number of sunspot groups observed over cycle 23 (in black) and those obtained in one realization of our model (in red) is shown in the left panel of Fig. 7. The similarity in the shape and spread of the two curves is quite evident. The same similarity is found when comparing the total group areas, i.e. the total area covered by sunspot groups in each day, (Fig. 7, right panel) and the group latitudes (Fig. 8) for the real and synthetic data.

### 4.2. Comparison test

In order to test and improve our model we have quantified how closely related the observed and synthetic data sets are by applying



**Fig. 8.** Same as in Fig. 7, but for the real and synthetic latitudinal distribution of sunspot groups.

ing the Kolmogorov-Smirnov test (Kolmogorov 1933; Smirnov 1939). This test compares the cumulative distribution functions of two samples, using the maximum deviation between them,

$$D_{n_1, n_2} = \max_x |F_{1, n_1}(x) - F_{2, n_2}(x)|, \quad (7)$$

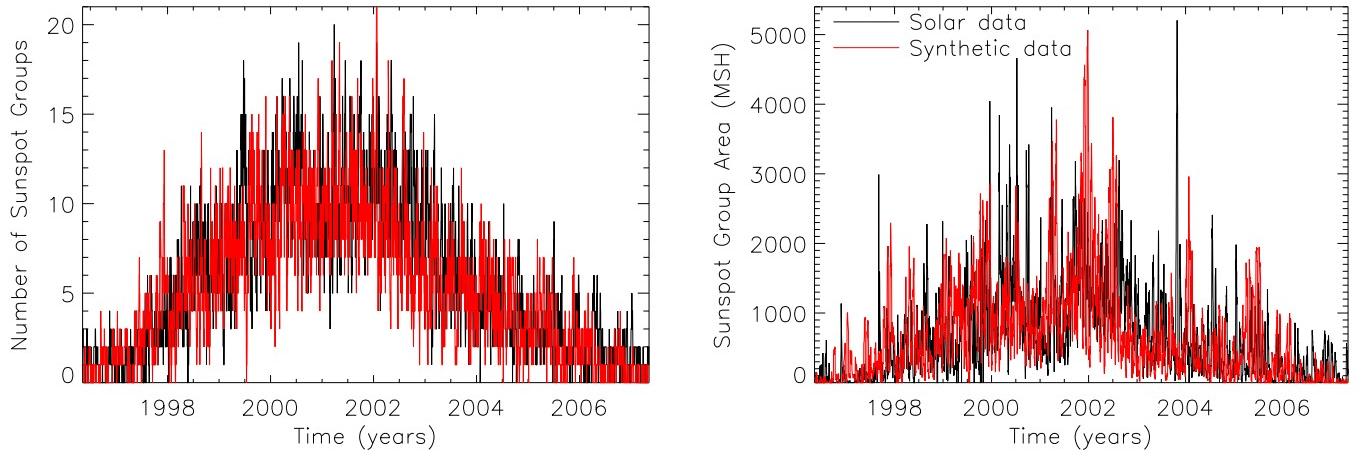
where  $n_1$  and  $n_2$  are the number of elements of sample 1 and sample 2, respectively, and  $F_1$  and  $F_2$  are the corresponding cumulative distribution functions.

The null hypothesis – i.e. that both samples result from the same distribution – is rejected at significance level  $\alpha$  if

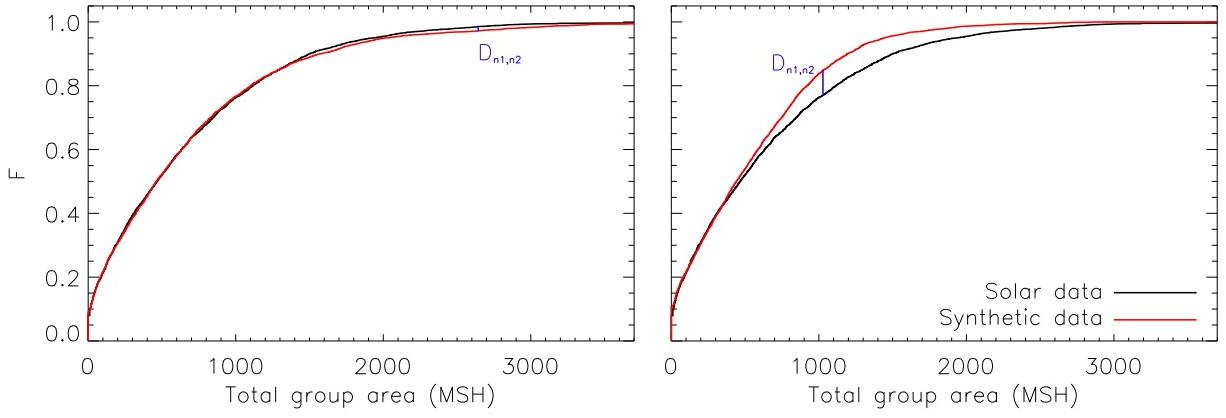
$$D_{n_1, n_2} > c(\alpha) \sqrt{\frac{n_1 + n_2}{n_1 n_2}}, \quad (8)$$

where  $c(\alpha)$  is a constant that depends of the significance level to be considered.

More than to reject the null hypothesis at a given level  $\alpha$ , in our case this test was used to identify the aspects of the model which needed to be improved. The synthetic sunspot cycles obtained from early versions of the model resulted in large values of  $D_{n_1, n_2}$ , indicating that they did not provide a good description of the observed properties of the sunspot cycle. By comparing the observed and synthetic cumulative distribution functions for the total area covered by sunspot groups and for the groups' latitudinal distribution, we could decide where and how to improve our model. An example of this is provided in Fig. 9, where we compare the cumulative distribution functions for the total group areas of the synthetic and real data. In this case,  $F$  corresponds to the fraction of days with total area below a given value, and  $n_1 = n_2 = 4017$  is the total number of days considered in the real and synthetic data. The left panel corresponds to a sunspot cycle reconstruction obtained with the current version of the model that considers an exponential function (equation (5)) to correct the GW rule for groups smaller than 85 MSH, while the right panel shows the results for a reconstruction obtained when adopting the GW rule for all group areas. Small groups from the former reconstruction live longer than groups with similar areas from the latter. This leads to an increase in the daily number of sunspot groups and, consequently, to a larger total area covered by the groups in the reconstruction obtained with the model that incorporates the corrected area-lifetime relation than with the other. The result is a shift of the cumulative distribution function towards larger areas and a better agreement with the observations.



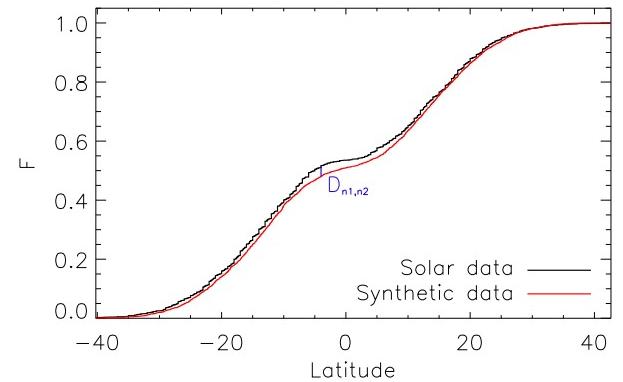
**Fig. 7.** Comparison between real solar data for cycle 23 (black) and synthetic data (red). Shown are the number of sunspot groups (left panel) and the total group area (right panel).



**Fig. 9.** Cumulative distribution functions of the synthetic (red) and real data (black) for the total group areas of a reconstruction obtained from the current version of the model (left panel;  $D_{n_1,n_2} \sim 0.0141$ ) and a reconstruction obtained when assuming the GW rule for the area-lifetime relation (right panel;  $D_{n_1,n_2} \sim 0.0798$ ).

Figure 10 compares the cumulative functions for the real and synthetic latitudes resulting from the same reconstruction, obtained with the current version of the empirical model. Here, the cumulative distribution functions,  $F$ , represent the fraction of groups that become visible (first appearance) at a latitude lower than a given value, and  $n_1 = 2801$  and  $n_2 = 2919$  correspond to the total number of different observed groups in the real and synthetic data, respectively. Despite the small value of the statistics  $D_{n_1,n_2}$ , the difference between the two distributions is significant and its interpretation is relatively straightforward: the cumulative distribution function for the solar data indicates that the southern hemisphere retains almost 55% of the visible sunspot groups, while for the synthetic data the groups are more evenly distributed by the two hemispheres. This discrepancy results from the hemispheric asymmetry that is known to be present in the data, but that is not accounted for in our model. To verify this, we compared the cumulative distribution functions for the absolute values of the latitude (Fig. 11), finding a lower value of the statistics  $D_{n_1,n_2}$ .

The results from the KS-test for the synthetic sunspot cycle discussed above are summarized in Table 1. At the significance level  $\alpha = 0.1$  the null hypothesis for the total group area and absolute group latitudes is not rejected. Although the non-rejection of the null hypothesis does not allow us to conclude about its veracity, it certainly reinforces the expectation born from the direct inspection of Figs. 7 and 8 that the synthetic cycles ob-



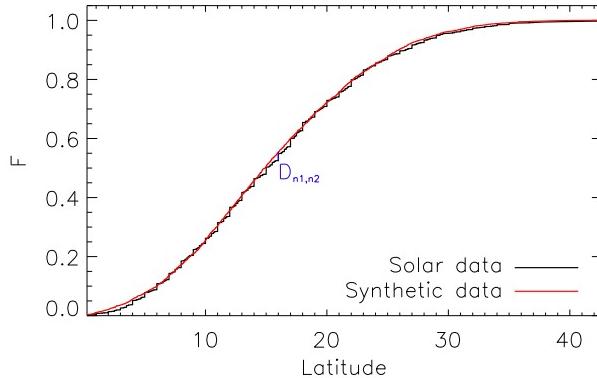
**Fig. 10.** Cumulative distribution functions of the synthetic (from the current model; red) and real data (black) for the spot latitudes. The maximum difference is  $D_{n_1,n_2} \sim 0.0377$  (blue).

tained from our model retain the main observed properties of the sunspot cycle.

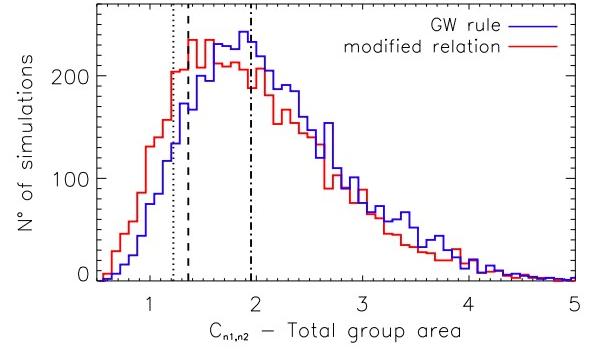
As the results from our empirical model are stochastic, we can perform Monte Carlo simulations to obtain the distributions for the statistics  $D_{n_1,n_2}$ . While  $n_2$  is constant for the total area covered by sunspot groups, for the latitudinal distribution  $n_2$  varies from reconstruction to reconstruction and accord-

	$n_1$	$n_2$	$D_{n_1,n_2}$	$\mathcal{D}(\alpha = 0.1)$
Total group area: Fig. 9, left	4017	4017	0.0141	0.0272
Latitudes - Fig. 10			0.0377	
Absolute Latitudes - Fig. 11	2801	2919	0.0283	0.0323

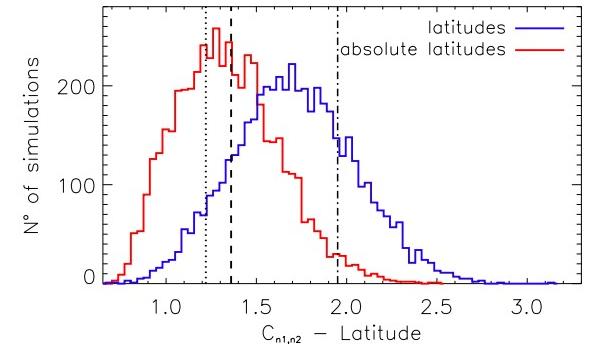
**Table 1.** Results from the KS-test for the synthetic sunspot cycle discussed in the text. The  $n_1$  and  $n_2$  are the sample sizes for the real and synthetic data, respectively.  $\mathcal{D} = c(0.1) \sqrt{(n_1 + n_2)/(n_1 n_2)}$  is the right-hand side of equation (8) for a significance  $\alpha = 0.1$ .



**Fig. 11.** Cumulative distribution functions of the synthetic (from the current model; red) and real data (black) for the absolute values of the group latitudes. The maximum deviation is  $D_{n_1,n_2} \sim 0.0283$ .



**Fig. 12.** Distribution of  $C_{n_1,n_2} = D_{n_1,n_2} / \sqrt{(n_1 + n_2)/(n_1 n_2)}$  for the total area covered by sunspot groups, when assuming the GW rule for all areas (blue histogram) and when using the modified area-lifetime relation (red histogram). The vertical lines indicate the levels of significance  $\alpha = 0.1$  (dotted),  $\alpha = 0.05$  (dashed) and  $\alpha = 0.001$  (dash-dotted).



**Fig. 13.** Distribution of  $C_{n_1,n_2} = D_{n_1,n_2} / \sqrt{(n_1 + n_2)/(n_1 n_2)}$  for the latitudinal distribution: from the group latitudes (blue) and absolute latitudes (red). The vertical lines indicate the levels of significance  $\alpha = 0.1$  (dotted),  $\alpha = 0.05$  (dashed) and  $\alpha = 0.001$  (dash-dotted).

ing to the case considered (both, southern or northern hemispheres). Thus, rather than considering the distributions for  $D_{n_1,n_2}$ , we consider those for the  $n_2$ -independent quantity  $C_{n_1,n_2} = D_{n_1,n_2} / \sqrt{(n_1 + n_2)/(n_1 n_2)}$ . Figures 12 and 13 summarize the results obtained from 5000 cycle reconstructions. The distribution of  $C_{n_1,n_2}$  obtained from the analysis of the total group area when the GW rule is assumed for all ranges of area (Fig. 12, blue histogram) is shifted towards larger values of  $C_{n_1,n_2}$  than that obtained from reconstructions that apply the correction to the area-lifetime relation for small groups (Fig. 12, red histogram). This confirms that the modified area-lifetime relation used in our model produces results that are statistically in better agreement with the solar data. Concerning the sunspot groups' latitudes (Fig. 13), the comparison of the distributions for  $C_{n_1,n_2}$  clearly confirms our earlier findings. The consequence of the non-inclusion of the hemisphere asymmetry in our model is that our synthetic cycles compare significantly better when considering the absolute values of the latitude.

The properties of the sunspot emergence are cycle-dependent, stronger cycles usually having higher latitudes and wider sunspot formation zones than weaker cycles (e.g. Solanki et al. 2008; Jiang et al. 2011). With this in mind we have applied our model to a second cycle, namely cycle 22, which is stronger and more asymmetric than cycle 23. In agreement with the works mentioned above, we found that the average latitude and the width of the sunspot formation zone obtained in our reconstructions for cycle 22 are larger than those obtained for cycle 23. Although the results from the KS-test for the latitudes were found to be worse for cycle 22 owing to the hemispheric asymmetry, we found that with regard to the total area covered by spots and to the absolute latitudes the synthetic data for this cycle is also in qualitative agreement with the real sunspot data.

## 5. Conclusions

In this work, we have presented an empirical tool for the stochastic reconstruction of sunspot cycles. With the parameters adopted in the version presented here, our tool produces synthetic daily sunspot records which retain the main properties of the real solar data.

A crucial assumption of our model is that different sunspot groups are generated independently. Despite evidence that sunspots tend to form within active longitudes (e.g. Bumba & Howard 1965; Bogart 1982; Jiang et al. 2011), pointing to possible correlations between their emergence, we found that the properties of the sunspot cycle are reasonably well reproduced under this model assumption. If a significant dependence between the generation existed we would expect that to have been reflected, for example, in a lack of agreement between the dispersion of the number of synthetic and observed sunspot groups as a function of time. In fact, we have checked that this problem

would occur if individual spots were (incorrectly) considered as independent events.

An important by-product of our work was the verification that the GW rule is only appropriate for the largest sunspot groups. For the smallest sunspot groups, this rule seems to underestimate the groups' lifetime. We have proposed a modified area-lifetime relation for small groups which leads to a closer agreement of the synthetic sunspot cycle with observations.

The quantity, quality, and diversity of solar data, can only be adequately reproduced by a relatively complex model that includes a number of empirical parameters and functions describing the average observed properties of the sunspot number, area, latitude and rotation. In contrast, for other stars the observational constraints are much more limited. Thus, the application of our tool to the study of activity cycles on other stars will require the identification of the model parameters that have a significant impact on the activity-related stellar observables, including the frequency shifts (work in progress). Another potential application of this kind of tool is related to the search for exoplanets, where new strategies for reducing activity signatures in the radial velocity and transit observations can be designed.

**Acknowledgements.** ARGS acknowledges the support from FCT (Portugal) through the grant reference SFRH/BD/88032/2012. MSC and PPA are supported by FCT through the Investigador FCT contracts of reference IF/00894/2012 and IF/00863/2012 and POPH/FSE (EC) by FEDER funding through the program COMPETE. Funds for this work were provided also by the EC, under FP7, through the projects FP7-SPACE-2012-31284 and PIRSES-GA-2010-269194. TLC acknowledges the support of the UK Science and Technology Facilities Council (STFC).

## References

- Abuzeid, B. & Marik, M. 1997, Sol. Phys., 173, 25
- Baglin, A., Michel, E., Auvergne, M., & Team, C. 2006, in ESA Special Publication, Vol. 624, Proceedings of SOHO 18/GONG 2006/HELAS I, Beyond the spherical Sun, 34
- Ballot, J., Gizon, L., Samadi, R., et al. 2011, A&A, 530, A97
- Baumann, I. & Solanki, S. K. 2005, A&A, 443, 1061
- Blanter, E. M., Le Mouël, J.-L., Perrier, F., & Shnirman, M. G. 2006, Sol. Phys., 237, 329
- Bogart, R. S. 1982, Sol. Phys., 76, 155
- Bogdan, T. J., Gilman, P. A., Lerche, I., & Howard, R. 1988, ApJ, 327, 451
- Borucki, W. J., Koch, D., Basri, G., et al. 2010, Science, 327, 977
- Brandt, P. N., Schmidt, W., & Steinegger, M. 1990, Sol. Phys., 129, 191
- Bravo, J. P., Roque, S., Estrela, R., Leão, I. C., & De Medeiros, J. R. 2014, A&A, 568, A34
- Bumba, V. & Howard, R. 1965, ApJ, 141, 1502
- Campante, T. L. 2012, PhD thesis, Universidade do Porto
- Campante, T. L., Handberg, R., Mathur, S., et al. 2011, A&A, A6
- Carrington, R. G. 1863, MNRAS, 23, 203
- Chang, H.-Y. 2012, NewA, 17, 247
- Chaplin, W. J., Elsworth, Y., Isaak, G. R., Miller, B. A., & New, R. 2004, MNRAS, 352, 1102
- Cincunegui, C., Díaz, R. F., & Mauas, P. J. D. 2007, A&A, 469, 309
- Czesla, S., Huber, K. F., Wolter, U., Schröter, S., & Schmitt, J. H. M. M. 2009, A&A, 505, 1277
- Du, Z. 2011, Sol. Phys., 273, 231
- Dumusque, X., Santos, N. C., Udry, S., Lovis, C., & Bonfils, X. 2011, A&A, 527, A82
- Duncan, D. K., Vaughan, A. H., Wilson, O. C., et al. 1991, ApJS, 76, 383
- Elling, W. & Schwentek, H. 1992, Sol. Phys., 137, 155
- Figueira, P., Marmier, M., Bonfils, X., et al. 2010, A&A, 513, L8
- García, R. A., Ceillier, T., Salabert, D., et al. 2014, ArXiv e-prints
- García, R. A., Mathur, S., Ballot, J., et al. 2010a, Science, 329, 1032
- García, R. A., Mathur, S., Salabert, D., et al. 2010b, arXiv:1008.4399
- Gleissberg, W. 1958, Z. Astrophys., 46, 219
- Gnevyshev, M. N. 1938, zv. Gl. Astron. Obs. Pulkove, 16, B36
- Gray, D. F. & Baliunas, S. L. 1995, ApJ, 441, 436
- Hale, G. E., Ellerman, F., Nicholson, S. B., & Joy, A. H. 1919, ApJ, 49, 153
- Hale, G. E. & Nicholson, S. B. 1925, ApJ, 62, 270
- Hall, J. C., Lockwood, G. W., & Skiff, B. A. 2007, ApJ, 133, 862
- Hathaway, D. H. 2010, Living Rev. Sol. Phys., 7, 1
- Hathaway, D. H. 2011, Sol. Phys., 273, 221
- Hathaway, D. H. & Choudhary, D. P. 2008, Sol. Phys., 250, 269
- Hathaway, D. H., Wilson, R. M., & Reichmann, E. J. 1994, Sol. Phys., 151, 177
- Henwood, R., Chapman, S. C., & Willis, D. M. 2009, Sol. Phys., 262, 299
- Howard, R. F. 1992, Sol. Phys., 137, 51
- Ivanov, V. G. & Miletzki, E. V. 2011, Sol. Phys., 268, 231
- Ivanov, V. G., Miletzki, E. V., & Nogovitsyn, Y. A. 2011, Astron. Rep., 55, 911
- Javaraiah, J. 2012, Astrophys. Space Sci., 338, 217
- Jiang, J., Cameron, R. H., Schmitt, D., & Schüssler, M. 2011, A&A, 528, A82
- Kiess, C., Rezaei, R., & Schmidt, W. 2014, A&A, 565, A52
- Kitchatinov, L. L. 2011, in Astronomical Society of India Conference Series, Vol. 2, Astronomical Society of India Conference Series, 71–80
- Koch, D. G., Borucki, W., Basri, G., et al. 2010, ApJ, 713, L79
- Kolmogorov, A. N. 1933, Giornale dell'Istituto Italiano degli Attuari, 4, 89
- Libbrecht, K. G. & Woodard, M. F. 1990, Nature, 345, 779
- Mathur, S., García, R. A., Ballot, J., et al. 2014, A&A, 562, A124
- Mathur, S., García, R. A., Catala, C., et al. 2010, A&A, 518, A53
- Maunder, E. W. 1904, MNRAS, 64, 747
- Metcalfe, T. S., Basu, S., Henry, T. J., et al. 2010, ApJ, 723, L213
- Metcalfe, T. S., Buccino, A. P., Brown, B. P., et al. 2013, ApJ, 763, L26
- Metcalfe, T. S., Dziembowski, W. A., Judde, P. G., & Snow, M. 2007, MNRAS, 379
- Miletzki, E. V. & Ivanov, V. G. 2009, Astron. Rep., 53, 857
- Moreno-Insertis, F. & Vázquez, M. 1988, A&A, 205, 289
- Mosser, B., Baudin, F., Lanza, A. F., et al. 2009, A&A, 506, 245
- Oshagh, M., Santos, N. C., Boisse, I., et al. 2013, A&A, 556, A19
- Oshagh, M., Santos, N. C., Ehrenreich, D., et al. 2014, A&A, 568, A99
- Petrovay, K. & van Driel-Gesztelyi, L. 1997, Sol. Phys., 166, 249
- Pont, F., Knutson, H., Gilliland, R. L., Moutou, C., & Charbonneau, D. 2008, MNRAS, 385, 109
- Saar, S. H. & Brandenburg, A. 1999, ApJ, 524, 295
- Sabarinath, A. & Anilkumar, A. K. 2008, Sol. Phys., 250, 183
- Schou, J., Antia, H. M., Basu, S., et al. 1998, ApJ, 505, 390
- Setiawan, J. Q. & Panofsky, A. A. 1938, ApJ, 88, 385
- Smirnov, N. V. 1939, Bulletin Moscow University, 2, 3
- Snodgrass, H. B. 1983, ApJ, 270, 288
- Snodgrass, H. B. & Ulrich, R. K. 1990, ApJ, 351, 309
- Solanki, S. K. 2003, A&ARv, 11, 153
- Solanki, S. K., Krivova, N. A., Schüssler, M., & Fligge, M. 2002, A&A, 396, 1029
- Solanki, S. K., Wenzler, T., & Schmitt, D. 2008, A&A, 483, 623
- Tripathy, S. C., Jain, K., Salabert, D., et al. 2011, Journal of Physics: Conference Series, 271, 012055:1
- Vaquero, M. J., Gordillo, A., Gallego, M. C., Sánchez-Bajo, F., & García, J. A. 2005, Observatory, 125, 152
- Volobuev, D. M. 2009, Sol. Phys., 258, 319
- Waldmeier, M. 1955, Ergebnisse und Probleme der Sonnenforschung (Leipzig: Geest & Portig)
- Wilson, O. C. 1978, ApJ, 226, 379
- Woodard, M. F. & Noyes, R. W. 1985, Nature, 318, 449
- Zappalà, R. A. & Zuccarello, F. 1991, A&A, 242, 480
- Zuccarello, F. 1993, A&A, 272, 587

## Appendix A: Schematic overview of the empirical model for the solar cycle

Figure A.1 provides a schematic summary of the procedure underlying our empirical tool. The green box illustrates the detailed treatment of the evolution of a sunspot group. All the parameters included in the model have been found to be important in order to obtain results that are in reasonable agreement with the observations. In the following scheme,  $U$ ,  $\text{Pois}$ ,  $\mathcal{N}$ , and  $\ln \mathcal{N}$  are, respectively, the uniform, Poisson, Gaussian, and log-normal distributions. The location in longitude of each sunspot group is represented by  $O$ .

